Multivariate Data Fusion and Uncertainty Quantification for Remote Sensing

Amy Braverman[1], Hai Nguyen[1], Noel Cressie[1,2], Anna Michalak[3], Emily Kang[4], Pulong Ma[4], Tim Stough[1], and Vineet Yadav[1]

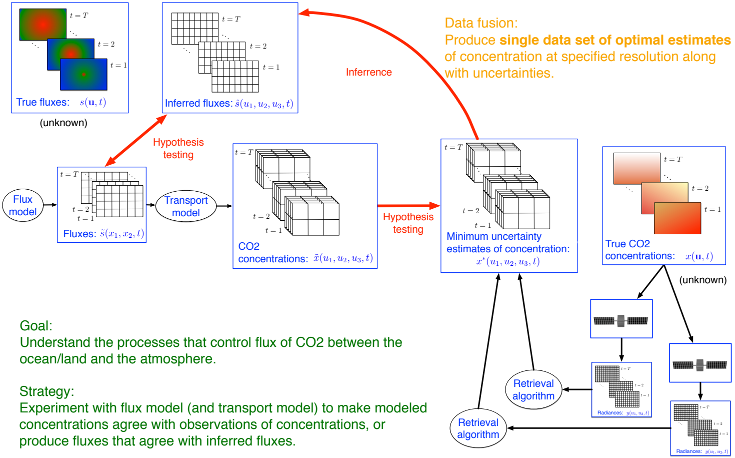
[1]Jet Propulsion Laboratory, California Institute of Technology
[2]National Institute of Applied Statistics Research Australia, University of Wollongong
[3]Carnegie Institution of Washington
[4]Department of Mathematical Sciences, University of Cincinnati

October 29, 2014

Outline

- Introduction and motivation.
- Mathematical/probabilistic framework.
- Modeling and exploiting spatial covariance.
- Modeling and exploiting temporal covariance.
- Fusing synthetic AIRS and OCO-2 profiles.
- How well did we do?
- Summary and conclusions.

Introduction and motivation



Goal:
Understand the processes that control flux of CO2 between the ocean/land and the atmosphere.

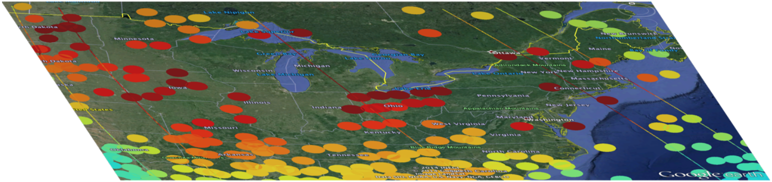
Strategy:
Experiment with flux model (and transport model) to make modeled concentrations agree with observations of concentrations, or produce fluxes that agree with inferred fluxes.

Introduction and motivation

- The goal of data fusion is to infer the values that make up a time-evolving spatial field from heterogeneous, noisy observations collected by multiple instruments.
- “Infer" = estimate the true value at any (or all) desired locations and times. Typically, this means on some grid at some pre-specified resolution.
- “Heterogeneous" = different footprints and sampling patterns.
- “Noisy" = different biases, measurement error variances, and missingness patterns.
- Exploit covariances in space, time, and among variables to make estimates with minimum uncertainty.

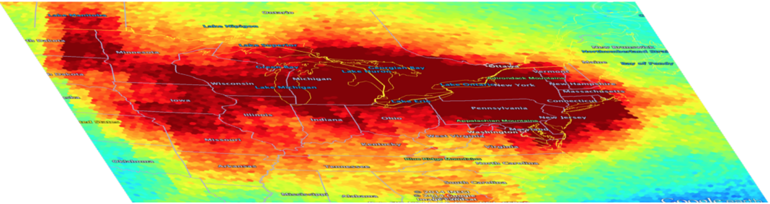
Introduction and motivation

Example: AIRS (circles) and OCO-2 (strips) synthetic data for a single time point:



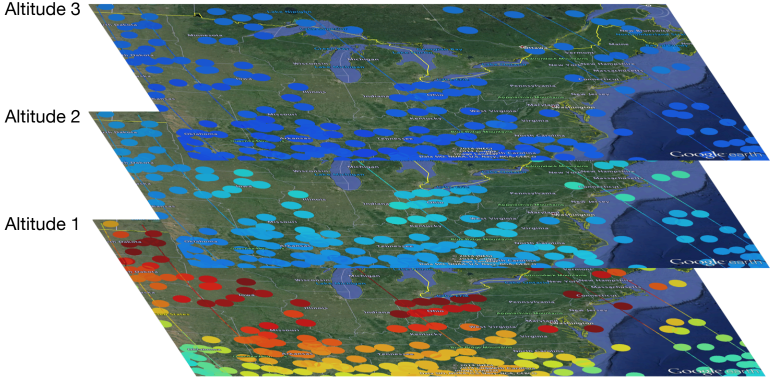
- AIRS footprints correspond to actual observed locations on January 1-3, 2006.
- OCO-2 footprints correspond to all possible observation locations (no filtering) for a single 3-day period (which one?).
- AIRS footprints = 90 km diameter. OCO-2 footprints ≈ 1 km footprints (strip = 4-across).

"True" (synthetic) field at at single time point:



- Find the estimate of the field that minimizes the uncertainty (estimate is unbiased and has minimum variance) by using all the OCO-2 and AIRS footprints to make estimates at all locations (and times!).

Introduction and motivation



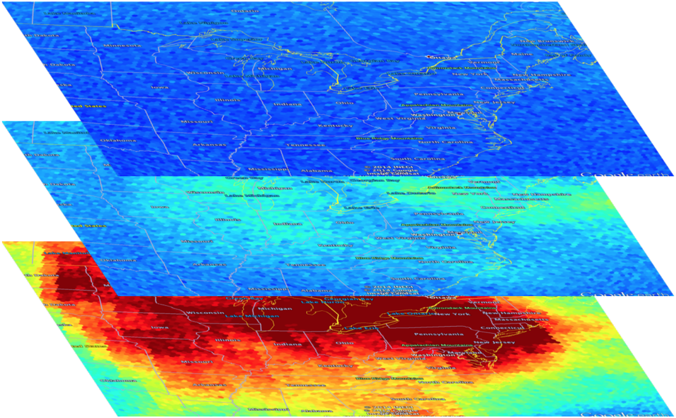
Multivariate data fusion: estimate vector-valued quantities, e.g., vertical profiles of CO2 mole-fraction.

Introduction and motivation

Altitude 3

Altitude 2

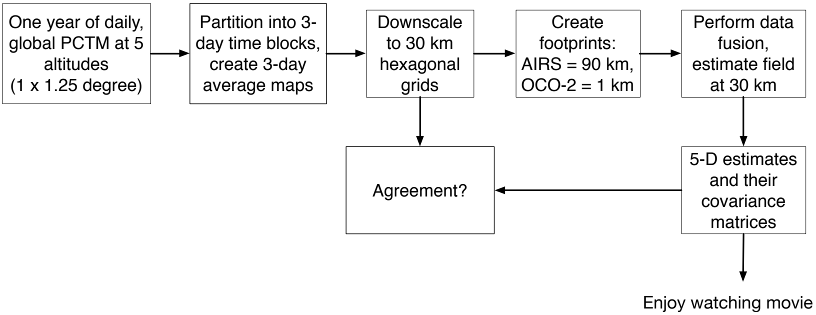
Altitude 1



- Find the minimum uncertainty estimate of the *multivariate* field using all the OCO-2 and AIRS observed profiles to make estimates at all locations, altitudes, and times.

Fusing synthetic AIRS and OCO-2 profiles

Fuse one year of synthetic AIRS and OCO-2 five-altitude profiles:

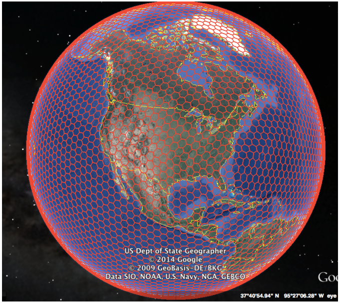


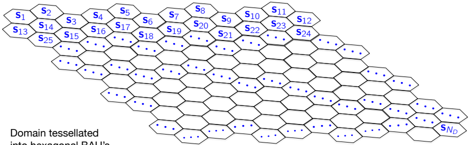
Fusing synthetic AIRS and OCO-2 profiles

Fused estimate, near surface CO2 mole-fraction (ppm):

Click me.

Mathematical/probabilistic framework





Domain tessellated into hexagonal BAU's

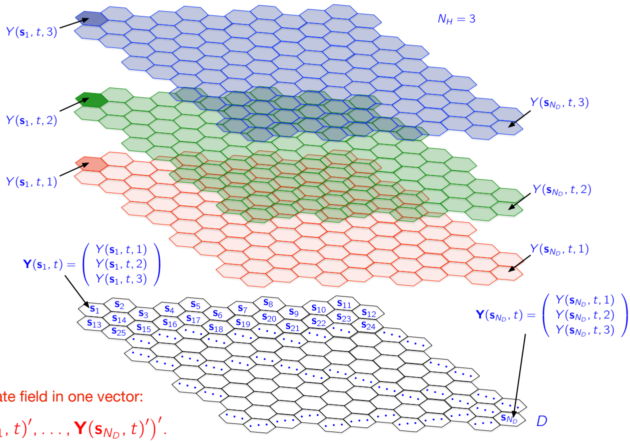
- Partition of Earth's surface into N_D (D is for "domain"), small hexagonal basic areal units (BAU's; 30 km in our application); the same at all time steps.
- BAU's indexed by **s**=lat/lon of their centers.
- Partition time into three-day blocks (basic time unit, BTU), indexed by t.
- At each BAU-BTU combination, there is a true but not directly observed vertical profile of CO2 mole-fraction,

$$\mathbf{Y}(\mathbf{s}, t) = (Y(\mathbf{s}, t, 1), \ldots, Y(\mathbf{s}, t, N_H))',$$

where N_H = number of altitudes.

Mathematical/probabilistic framework

Geophysical field:

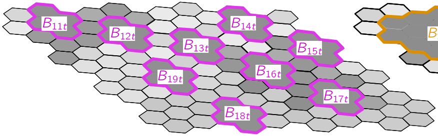


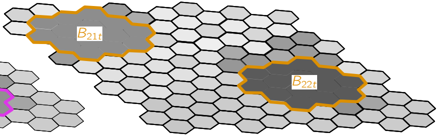
The whole multivariate field in one vector:

$$\mathbf{Y}_t = \left(\mathbf{Y}(\mathbf{s}_1, t)', \ldots, \mathbf{Y}(\mathbf{s}_{N_D}, t)'\right)'.$$

Mathematical/probabilistic framework

Observations are the averages of BAU values within instrument footprints, plus footprint-level measurement error.





$$\mathbf{Z}^{(1)}(B_{1it}) = \frac{1}{|D \cap B_{1it}|} \sum_{\mathbf{s} \in D \cap B_{1it}} \mathbf{Y}(\mathbf{s}, t) + \boldsymbol{\epsilon}(B_{1it})$$

$$\mathbf{Z}^{(2)}(B_{2jt}) = \frac{1}{|D \cap B_{2jt}|} \sum_{\mathbf{s} \in D \cap B_{2jt}} \mathbf{Y}(\mathbf{s}, t) + \boldsymbol{\epsilon}(B_{2jt})$$

All instrument 1 observations in one vector:

$$\mathbf{Z}_t^{(1)} = \left(\mathbf{Z}^{(1)}(B_{11t})', \ldots, \mathbf{Z}^{(1)}(B_{1N_t^{(1)}t})'\right)'$$

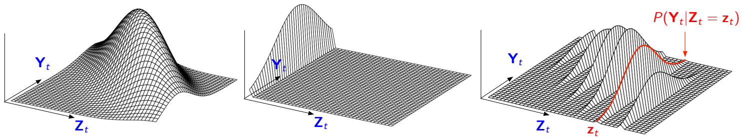
All instrument 2 observations in one vector:

$$\mathbf{Z}_t^{(2)} = \left(\mathbf{Z}^{(2)}(B_{21t})', \ldots, \mathbf{Z}^{(2)}(B_{1N_t^{(2)}t})'\right)'$$

All observations in one vector: $\mathbf{Z}_t = \left(\mathbf{Z}_t^{(1)\prime}, \mathbf{Z}_t^{(2)\prime}\right)'$.

Mathematical/probabilistic framework

- $\mathbf{Z}_t$ is the vector of all “noisy" observations (measurement and aggregation error).
- $\mathbf{Y}_t$ is the vector of all unknown (uncertain and not directly observed) values of the high-resolution spatial field.
- We want to estimate $\mathbf{Y}_t$ given $\mathbf{Z}_t$.

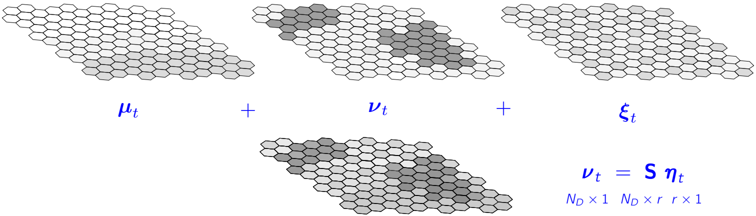


- The minimum uncertainty (unbiased, minimum variance) estimate of $\mathbf{Y}_t$ given the observed data, $\mathbf{Z}_t$, is $\mathrm{E}(\mathbf{Y}_t|\mathbf{Z}_t)$. The uncertainty is $\mathrm{var}(\mathbf{Y}_t|\mathbf{Z}_t)$. (Expected value and covariance matrix of the posterior distribution of $\mathbf{Y}_t$ given $\mathbf{Z}_t$.)

Modeling and exploiting spatial covariance

Strategy: break $\mathbf{Y}_t$ into pieces, estimate pieces separately.

trend | small-scale variation | fine-scale variation



$$\boldsymbol{\mu}_t \quad + \quad \boldsymbol{\nu}_t \quad + \quad \boldsymbol{\xi}_t$$

$$\mathbf{Y}_t$$

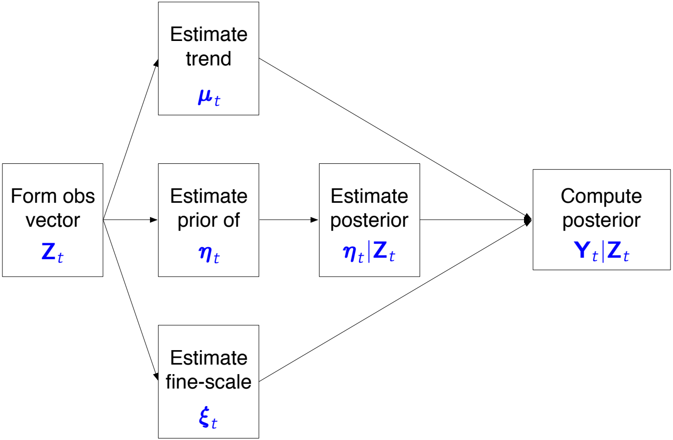
$$\underset{N_D \times 1}{\boldsymbol{\nu}_t} = \underset{N_D \times r}{\mathbf{S}} \; \underset{r \times 1}{\boldsymbol{\eta}_t}$$

$\mathbf{S}$ = matrix of spatial basis functions
$\boldsymbol{\eta}_t$ = coefficients

The field $\mathbf{Y}_t$ is the super-position of three independent components: the trend, $\boldsymbol{\mu}_t$, the small-scale variation, $\boldsymbol{\nu}_t$, and the fine-scale variation, $\boldsymbol{\xi}_t$. Write

$$\mathbf{Y}_t = \boldsymbol{\mu}_t + \boldsymbol{\nu}_t + \boldsymbol{\xi}_t, \quad \text{and} \quad \mathrm{E}(\mathbf{Y}_t|\mathbf{Z}_t) = \mathrm{E}(\boldsymbol{\mu}_t|\mathbf{Z}_t) + \mathbf{S}\,\mathrm{E}(\boldsymbol{\eta}_t|\mathbf{Z}_t) + \mathrm{E}(\boldsymbol{\xi}_t|\mathbf{Z}_t),$$
$$\mathrm{cov}(\mathbf{Y}_t|\mathbf{Z}_t) = \mathrm{cov}(\boldsymbol{\mu}_t|\mathbf{Z}_t) + \mathbf{S}\,\mathrm{cov}(\boldsymbol{\eta}_t|\mathbf{Z}_t)\,\mathbf{S}' + \mathrm{cov}(\boldsymbol{\xi}_t|\mathbf{Z}_t).$$

Modeling and exploiting spatial covariance

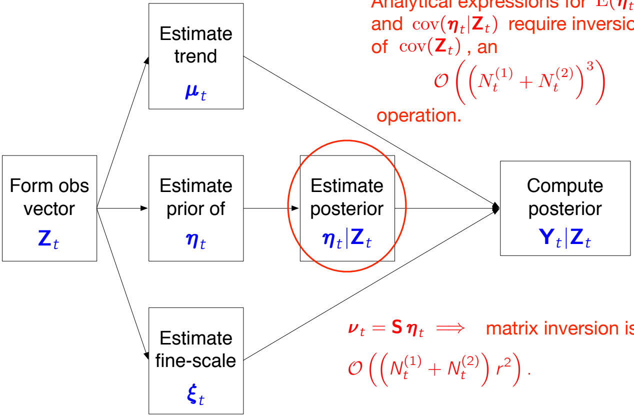


Modeling and exploiting spatial covariance

Analytical expressions for $\mathrm{E}(\boldsymbol{\eta}_t|\mathbf{Z}_t)$ and $\mathrm{cov}(\boldsymbol{\eta}_t|\mathbf{Z}_t)$ require inversion of $\mathrm{cov}(\mathbf{Z}_t)$, an

$$\mathcal{O}\left(\left(N_t^{(1)} + N_t^{(2)}\right)^3\right)$$

operation.



$\boldsymbol{\nu}_t = \mathbf{S}\,\boldsymbol{\eta}_t \implies$ matrix inversion is

$$\mathcal{O}\left(\left(N_t^{(1)} + N_t^{(2)}\right) r^2\right).$$

Modeling and exploiting temporal covariance

Apply Kalman Smoother to $\boldsymbol{\eta}_t$ (Nguyen et al., 2013):

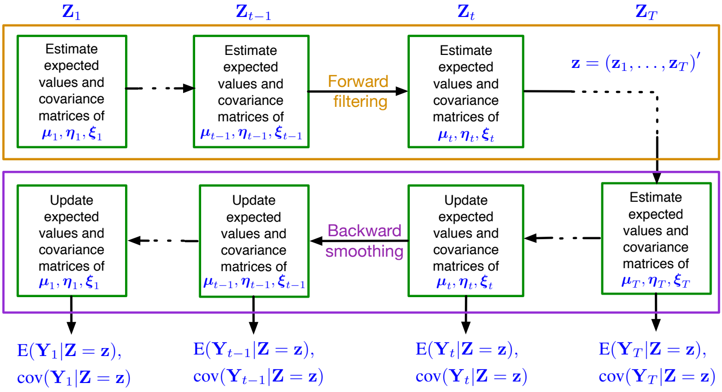
- Model the temporal evolution of $\boldsymbol{\eta}_t$ as an auto-regressive process:

$$\boldsymbol{\eta}_{t+1} = \mathbf{H}\boldsymbol{\eta}_t + \boldsymbol{\zeta}_t, \quad \boldsymbol{\zeta}_t \sim N(\mathbf{0}, \mathbf{U}),$$

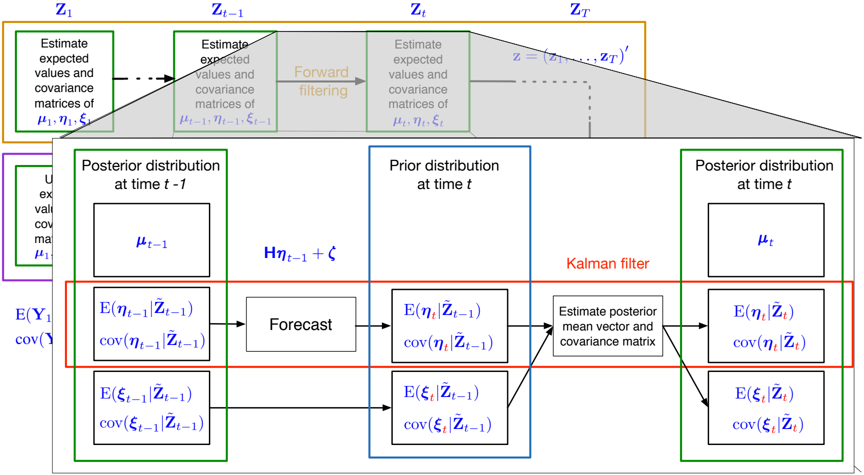
  where $\mathbf{H}$ is the "propagator" matrix, and $\boldsymbol{\zeta}_t$ is the "innovation" matrix.

- Estimate $\mathbf{H}$ and $\mathbf{U}$ from the observations.
- Forward filtering: for each time block (BTU) $t = 1, \ldots, T$, obtain maximum likelihood estimates (via the EM algorithm) of the parameters of posterior distribution of $\boldsymbol{\eta}_t$.
- Backward smoothing: for each time block, filter backwards in time so that the estimates are based on *all* data from all time blocks.

Modeling and exploiting temporal covariance



Modeling and exploiting temporal covariance



$$\tilde{\mathbf{Z}}_t = (\mathbf{Z}_t, \mathbf{Z}_{t-1}, \ldots, \mathbf{Z}_1)\,.$$

Fusing synthetic AIRS and OCO-2 profiles

Fuse one year of synthetic AIRS and OCO-2 five-altitude profiles:

- Synthetic truth field (five altitudes) created by downscaling output of the Parameterized Chemistry Transport Model (PCTM).
  - 365 daily model runs at $1^\circ \times 1.25^\circ$ resolution.
  - Downscaled to 30 km resolution using conditional simulation (Stough et al., 2014).
- Synthetic AIRS footprints (90 km) obtained by averaging 30 km hexagons belonging to actual AIRS footprints for corresponding day of 2006 (cloud-screened).
- Synthetic OCO-2 footprints ($\approx$ 1 km) obtained as value of 30 km hexagon to which footprint center belongs for representative orbit tracks (not cloud-screened).
- No measurement error (yet).

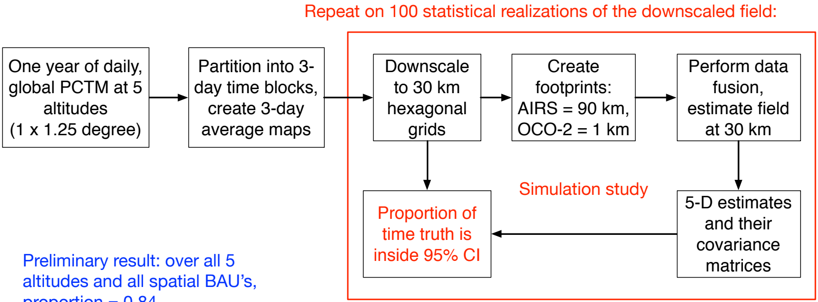
Fusing synthetic AIRS and OCO-2 profiles

Fuse one year of synthetic AIRS and OCO-2 five-altitude profiles:

- Time aggregated into three-day blocks, Kalman smoother run on monthly "windows" (ten three-day blocks per month). Propagator matrix and innovation vector re-estimated for each window.
- About 40,000 AIRS and 200,000 OCO-2 synthetic observations per three-day block.
- We used $r \approx 1800$ basis centers in three dimensions (300 horizontal $\times$ 6 vertical at each horizontal location).
- Estimated five-altitude profile and their covariance matrices produced at 30 km BAU resolution globally for 120, three-day time blocks covering one (synthetic) year.
- Timing: fusing one month (five altitudes) in ten, three-day blocks takes about 36 hours on a single Intel Xeon 2.0 Ghz processor.

How well did we do?

Fuse one year of synthetic AIRS and OCO-2 five-altitude profiles:



Preliminary result: over all 5 altitudes and all spatial BAU's, proportion = 0.84.

Summary and conclusions

- Spatial (and inter-variable) dependence captured by a combination of basis functions and a low-dimensional hidden state vector. Estimation performed in low-dimensional space. No assumptions of isotropy or stationarity required.
- Temporal dependence via a Kalman smoother on the hidden state.
- Corrects for change of support (heterogenous footprints) and different measurement error characteristics.
- Computationally feasible for very large remote sensing data sets.
- *No instrument observes everywhere all the time, or perfectly. Here we leverage complementary strengths of multiple instruments to increase coverage and minimize uncertainty.*

Summary and conclusions

- Still work to do in evaluating results through simulation studies.
- Still work to do on the selection of basis functions and interplay between them, the trend, and the fine-scale term.
- Preparing to apply to actual AIRS and OCO-2 data early next year.
- Journal paper in preparation.

References

Nguyen, H., Cressie, N., and Braverman, A. (2012). Spatial Statistical Data Fusion for Remote- Sensing Applications, *Journal of the American Statistical Association*, 107, pp. 1004-1018.

Nguyen, H., Katzfuss, M., Cressie, N., and Braverman, A. (2013). Spatio-Temporal Data Fusion for Very Large Remote Sensing Datasets, *Technometrics*, DOI: 10.1080/00401706.2013.831774.

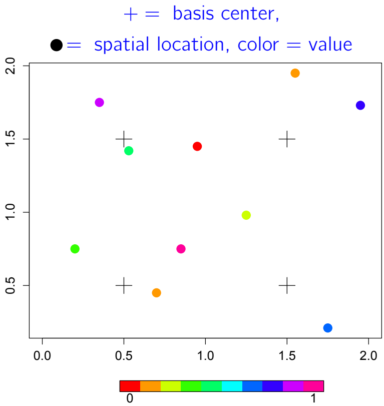
This work was partially performed at the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration. Government sponsorship acknowledged.

Research described here is made possible by NASA's Earth Science Technology Office through its Advanced Information Systems Technology (AIST-11) program.

Backup slides

Spatial basis functions

2-D cartoon example, 10 locations, 4 basis centers:



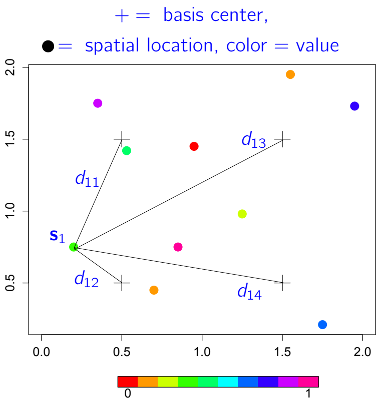
Spatial structure given by $\mathrm{cov}(\boldsymbol{\nu}_t)$.

10×10

$$\boldsymbol{\nu}_t = (\nu(\mathbf{s}_1, t), \ldots, \nu(\mathbf{s}_{10}, t))'$$

Spatial basis functions

2-D cartoon example, 10 locations, 4 basis centers:



Basis function for each location is a decaying function of its distance to the four basis centers:

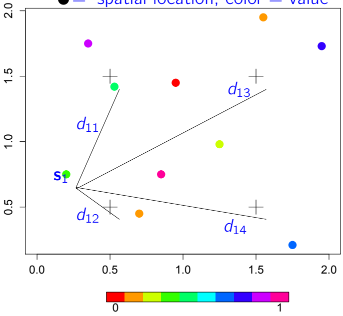
$$\mathbf{S}(\mathbf{s}_1) = (1/d_{11}, 1/d_{12}, 1/d_{13}, 1/d_{14})\,.$$

Spatial basis functions

2-D cartoon example, 10 locations, 4 basis centers:

$+ =$ basis center,

$\bullet =$ spatial location, color $=$ value



$\boldsymbol{\nu}_t$

Basis function matrix:

$$\mathbf{S} = \begin{pmatrix} \mathbf{S}(\mathbf{s}_1) \\ \mathbf{S}(\mathbf{s}_2) \\ \vdots \\ \mathbf{S}(\mathbf{s}_{10}) \end{pmatrix} = \begin{pmatrix} 1/d_{11} & 1/d_{12} & 1/d_{13} & 1/d_{14} \\ 1/d_{21} & 1/d_{22} & 1/d_{23} & 1/d_{24} \\ \vdots & \vdots & \vdots & \vdots \\ 1/d_{10,1} & 1/d_{10,2} & 1/d_{10,3} & 1/d_{10,4} \end{pmatrix}$$

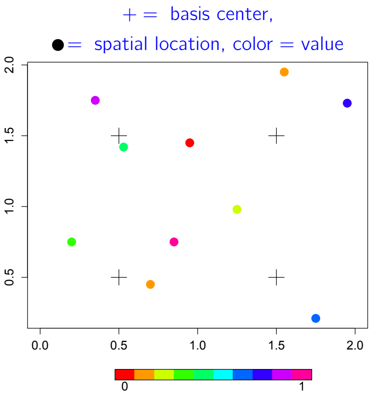
Low-dimensional representation:

$$\mathbf{S}\,\boldsymbol{\eta}_t = \begin{pmatrix} 1/d_{11} & 1/d_{12} & 1/d_{13} & 1/d_{14} \\ 1/d_{21} & 1/d_{22} & 1/d_{23} & 1/d_{24} \\ \vdots & \vdots & \vdots & \vdots \\ 1/d_{10,1} & 1/d_{10,2} & 1/d_{10,3} & 1/d_{10,4} \end{pmatrix} \begin{pmatrix} \eta_{1t} \\ \eta_{2t} \\ \eta_{3t} \\ \eta_{4t} \end{pmatrix}$$

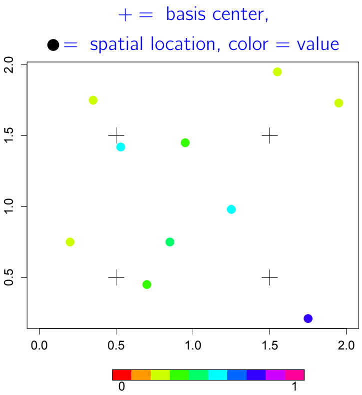
$$\mathrm{cov}(\boldsymbol{\nu}_t) = \mathrm{cov}(\mathbf{S}\,\boldsymbol{\eta}_t) = \mathbf{S}\,\underbrace{\mathrm{cov}(\boldsymbol{\eta}_t)}_{4 \times 4}\,\mathbf{S}'$$

Spatial basis functions

2-D cartoon example, 10 locations, 4 basis centers:



$\boldsymbol{\nu}_t$



$\hat{\boldsymbol{\nu}}_t = \mathbf{S}\,\hat{\boldsymbol{\eta}}_t.$

Modeling and exploiting spatial covariance

- Have $P(\mathbf{Z}_t|\boldsymbol{\eta}_t)$, want $P(\boldsymbol{\eta}_t|\mathbf{Z}_t)$. Use Bayes' Theorem $(P(B|A) \propto P(A|B)P(B).)$

$$\mathbf{Y}(\mathbf{s}, t) = \boldsymbol{\mu}(\mathbf{s}, t) + \boldsymbol{\nu}(\mathbf{s}, t) + \boldsymbol{\xi}(\mathbf{s}, t)$$

$$\swarrow \qquad \searrow$$

$$\mathbf{Z}^{(1)}(B_{1it}) = \frac{1}{|D \cap B_{1it}|} \sum_{s \in B_{1it}} \mathbf{Y}(\mathbf{s}, t) + \boldsymbol{\epsilon}(B_{1it}) \qquad \mathbf{Z}^{(2)}(B_{2jt}) = \frac{1}{|D \cap B_{2jt}|} \sum_{s \in B_{2jt}} \mathbf{Y}(\mathbf{s}, t) + \boldsymbol{\epsilon}(B_{2jt})$$

$$\mathbf{Z}_t^{(1)} = \boldsymbol{\mu}_t^{(1)} + \mathbf{S}^{(1)}\boldsymbol{\eta}_t^{(1)} + \boldsymbol{\xi}_t^{(1)} + \boldsymbol{\epsilon}_t^{(1)} \qquad \mathbf{Z}_t^{(2)} = \boldsymbol{\mu}_t^{(2)} + \mathbf{S}^{(2)}\boldsymbol{\eta}_t^{(2)} + \boldsymbol{\xi}_t^{(2)} + \boldsymbol{\epsilon}_t^{(2)}$$

$$\searrow \qquad \swarrow$$

$$\begin{pmatrix} \mathbf{Z}_t^{(1)} \\ \mathbf{Z}_t^{(2)} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\mu}_t^{(1)} \\ \boldsymbol{\mu}_t^{(2)} \end{pmatrix} + \begin{pmatrix} \mathbf{S}_t^{(1)} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_t^{(2)} \end{pmatrix} \begin{pmatrix} \boldsymbol{\eta}_t^{(1)} \\ \boldsymbol{\eta}_t^{(2)} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\xi}_t^{(1)} \\ \boldsymbol{\xi}_t^{(2)} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\epsilon}_t^{(1)} \\ \boldsymbol{\epsilon}_t^{(2)} \end{pmatrix}$$

$$\Downarrow$$

$$\mathbf{Z}_t = \boldsymbol{\mu}_t + \mathbf{S}\,\boldsymbol{\eta}_t + \boldsymbol{\xi}_t + \boldsymbol{\epsilon}_t$$

- $P(\boldsymbol{\eta}_t|\mathbf{Z}_t) \propto P(\mathbf{Z}_t|\boldsymbol{\eta}_t)P(\boldsymbol{\eta}_t)$.

Vertical basis functions

- In previous work (Nguyen, Katzfuss, Cressie, and Braverman (2012)) we used 446 basis centers arranged in a multi-resolution configuration with local bisquare decay to capture 2-D spatial structure in $\boldsymbol{\nu}_t$.

- Basis functions for 3-D location$(\mathbf{s}, h)$ is $\mathcal{S}(\mathbf{s}, h)$. It is the Kronecker product of the horizontal basis function, $\mathbf{S}(\mathbf{s})$, and vertical (horizontally varying) basis function $\boldsymbol{\tau}(\mathbf{s}, h)$:

$$\mathcal{S}(\mathbf{s}, h) = \mathbf{S}(\mathbf{s}) \otimes \boldsymbol{\tau}(\mathbf{s}, h).$$

Example:

$$\mathbf{S}(\mathbf{s}) = \begin{pmatrix} S_1 \\ \vdots \\ S_{r_1} \end{pmatrix}, \quad \boldsymbol{\tau}(\mathbf{s}, h) = \begin{pmatrix} \tau_1 \\ \vdots \\ \tau_{r_2} \end{pmatrix}, \quad \mathbf{S}(\mathbf{s}) \otimes \boldsymbol{\tau}(\mathbf{s}, h) = \begin{pmatrix} S_1 \tau_1 \\ S_1 \tau_2 \\ \vdots \\ S_1 \tau_{r_2} \\ \vdots \\ S_{r_1} \tau_1 \\ S_{r_1} \tau_2 \\ \vdots \\ S_{r_1} \tau_{r_2} \end{pmatrix}.$$

- $\boldsymbol{\tau}(\mathbf{s}, h)$ expands h from one number to a vector of six numbers in a way that depends on location $\mathbf{s}$.

Modeling and exploiting spatial covariance

The data model relates each instrument footprint observed value to the true process:

$$\mathbf{Z}_t^{(k)} = \begin{pmatrix} \mathbf{Z}^{(k)}(B_{k1t}) \\ \vdots \\ \mathbf{Z}^{(k)}(B_{kN_t^{(k)}t}) \end{pmatrix}, \quad \mathbf{Z}^{(k)}(B_{kit}) = \mathbf{Y}^{(k)}(B_{kit}, t) + \epsilon(B_{kit}),$$

$$\mathbf{Y}^{(k)}(B_{kit}) = \left[\frac{1}{|D \cap B_{kit}|} \sum_{\mathbf{s} \in |D \cap B_{kit}|} \mathbf{Y}(\mathbf{s}, t) \right] \quad \text{(noiseless spatial aggregate)},$$

$$= \left[\frac{1}{|D \cap B_{kit}|} \sum_{\mathbf{s} \in |D \cap B_{kit}|} \boldsymbol{\mu}^{(k)}(\mathbf{s}, t) + \mathbf{S}(\mathbf{s})\boldsymbol{\eta}_t + \boldsymbol{\xi}^{(k)}(\mathbf{s}, t) \right].$$